Estimating percent surface-water area using intermediate resolution satellite imagery

Jennifer Rover[1], Bruce Wylie[1], and Lei Ji[2]

[1]U.S. Geological Survey (USGS) Earth Resources Observation and Science (EROS) Center, Sioux Falls, SD, USA.
[2]ASRC Research and Technology Solutions, contractor to the USGS EROS Center, Sioux Falls, SD, USA. Work performed under USGS contract 08HQCN0007.

Introduction

Multi-spectral, intermediate spatial resolution satellite data, such as Landsat TM/ETM+, have been used widely for mapping surface-water bodies at regional and national scales. Accurate estimation of surface-water area, however, still remains a challenge because the intermediate resolution images are not capable of detecting small wetlands or small changes in wetlands that are of great ecological significance (Downing et al., 2006; Krankina et al., 2008). To compensate for the limitations of the intermediate resolution images for mapping small water bodies, a fuzzy classification method can be used to estimate subpixel fractions of water and produce a map of continuous percent water area. Such methods usually require a large number of field training sites or pairs of moderate and high resolution images from the same time period. We avoid these limitations by developing a regression-based fuzzy classification technique capable of collecting training data from the Landsat image itself to map water features. We applied the method to study sites in the Yukon Flats of Alaska and the Prairie Pothole Region of North Dakota. Three other statistical classification techniques were also applied to the same Yukon Flats image for comparison.

Water Mapping Methods

The initial process included producing water and nonwater maps from a Landsat image (30-m resolution) and a SPOT image (5-m and 10-m resolution in the Yukon Flats and PPR, respectively) acquired on corresponding dates. The binary water maps were generated with a See5 decision tree model (Quinlan, 1993, 2005) that captured surface water at both resolutions with error rates of 1.5% (SPOT) and 3.3% (Landsat). Predictor variables included the spectral bands, NDVI (Normalized Difference Vegetation Index), NDPI (Normalized Difference Pond Index), band 4 and 5 ratios, principal components, and canopy cover (LANDFIRE).
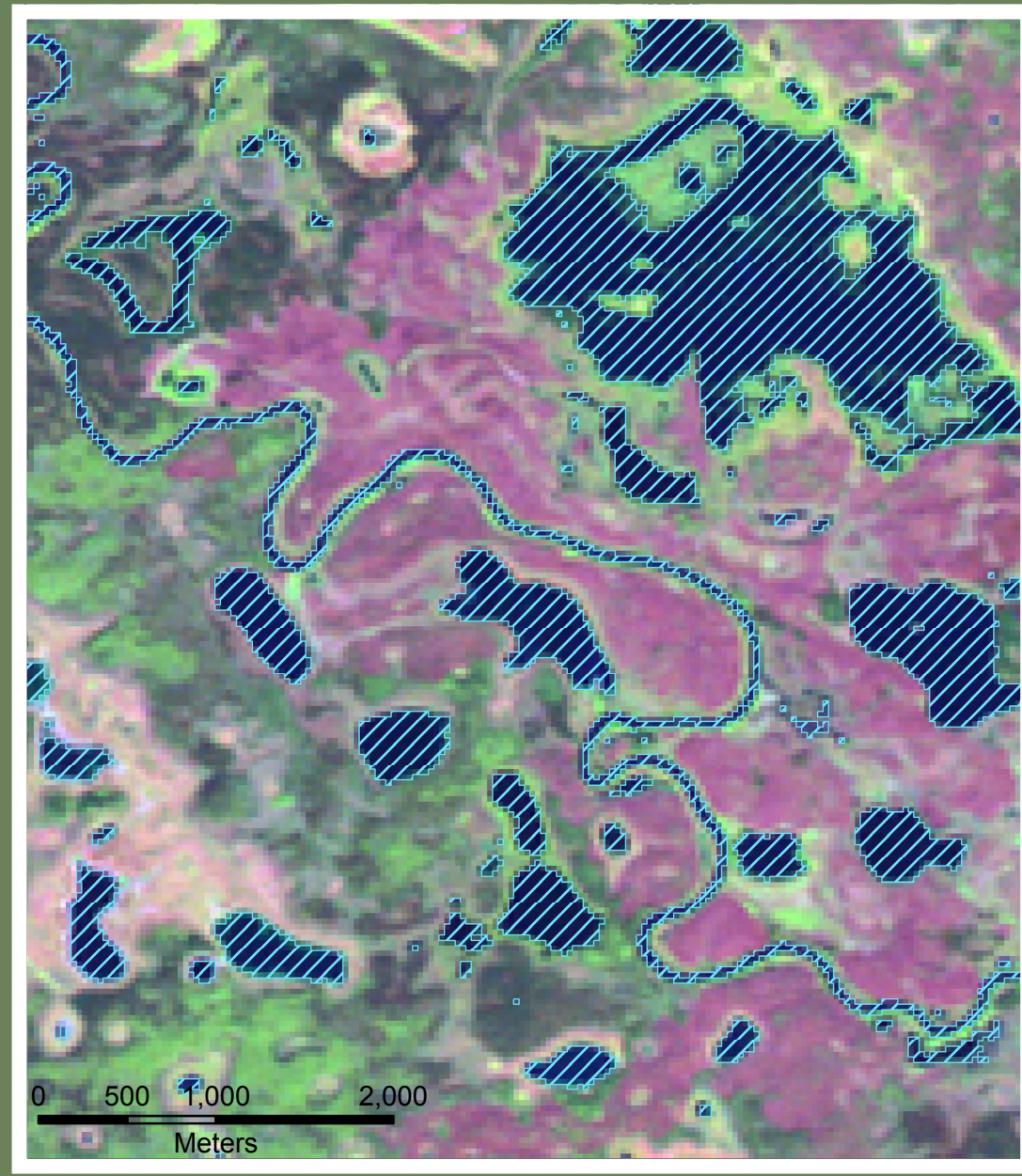
Landsat TM (RGB = 5,4,3) and water (blue hatch) on August 20, 2007 in the lower mouth of Birch Creek and Canvasback Lake, in the Yukon Flats of Alaska, developed from a decision tree model.
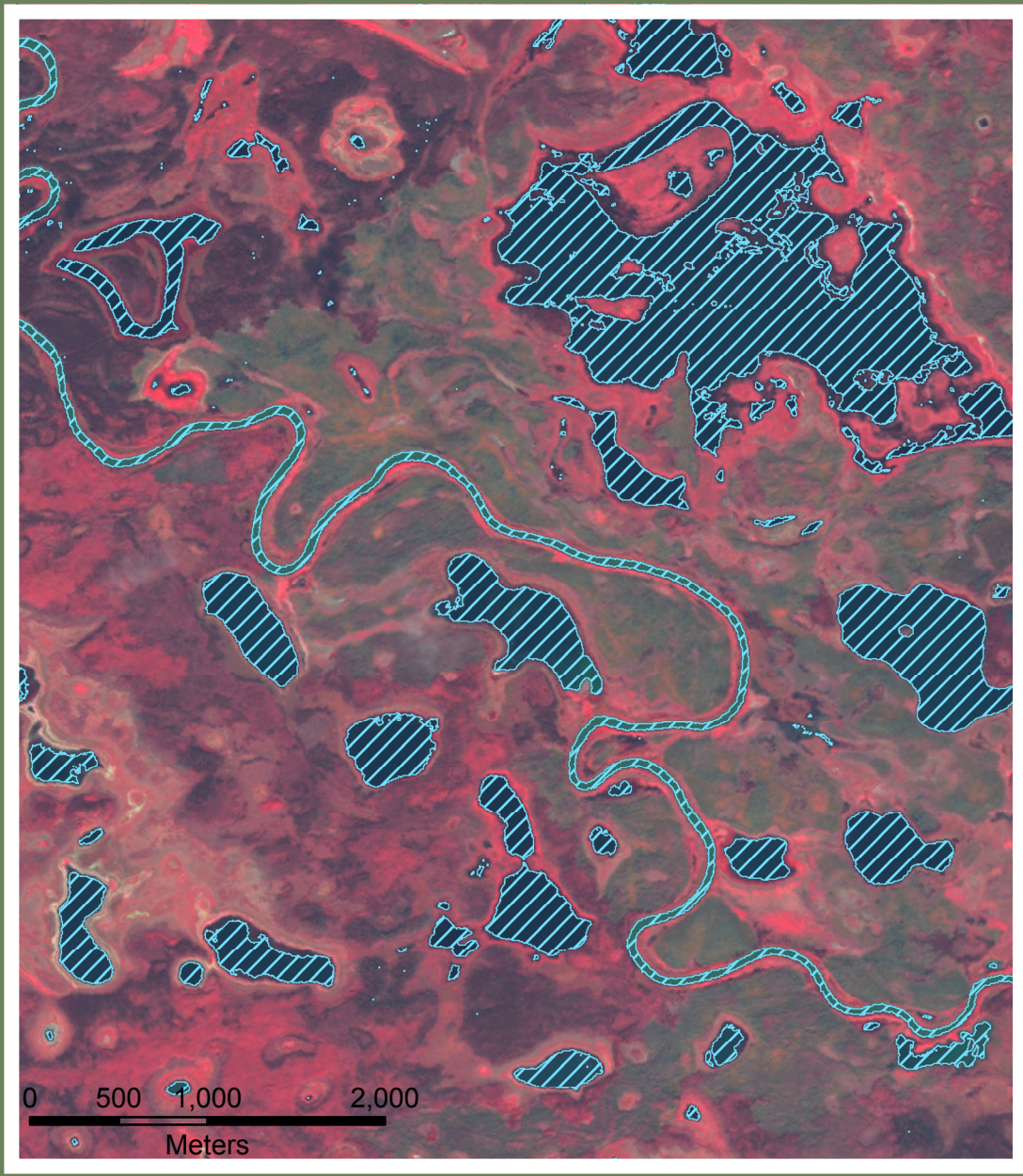
SPOT 5 (RGB = 1,2,3) and water (blue hatch) on August 28, 2007 in the lower mouth of Birch Creek and Canvasback Lake, in the Yukon Flats of Alaska, developed from a decision tree model.

Percent Water Methods, Results, and Validation

The water extents derived from the 5-m SPOT image, acquired at an analogous date, were used to determine percent water within each Landsat 30-m pixel. The SPOT 30-m and the Landsat 30-m subpixel percent water images were then scaled to 90 m for a final assessment of the Landsat-estimated percent water area products. The new method produced a model with a higher R^2 value (0.92) than the endmember (R^2=0.89) or linear unmixing classification (R^2=0.86) methods. The new method was comparable to our control, percent water mapped directly from a SPOT image (R^2=0.93), and does not require high resolution images from the same time period for model development. For surface water change studies, the new method provides an opportunity to utilize archival moderate resolution imagery.

A comparison of four statistical classification approaches to evaluate a new method for producing 30-m subpixel percent water maps from Landsat imagery

Our new method produced subpixel percent water from Landsat data. The approach used a regression tree model where the predictor variables are the spatial average of a 3- by 3-pixel (90- by 90-m) window for each Landsat spectral reflectance band and several derivatives (e.g., NDVI). The response variable, percent water, was calculated from the number of water pixels in the TM water image that were located within the same 3- by 3-pixel (90- by 90-m) window area. The regression tree model, developed from the percent water estimated within each 90-m window, is then applied to the 30-m Landsat image.
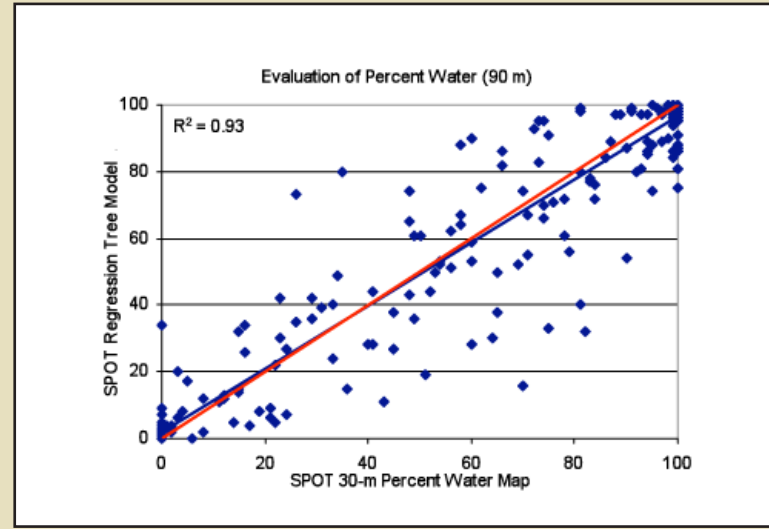
**Self-trained Regression Tree Model (MSE: 13.07)**
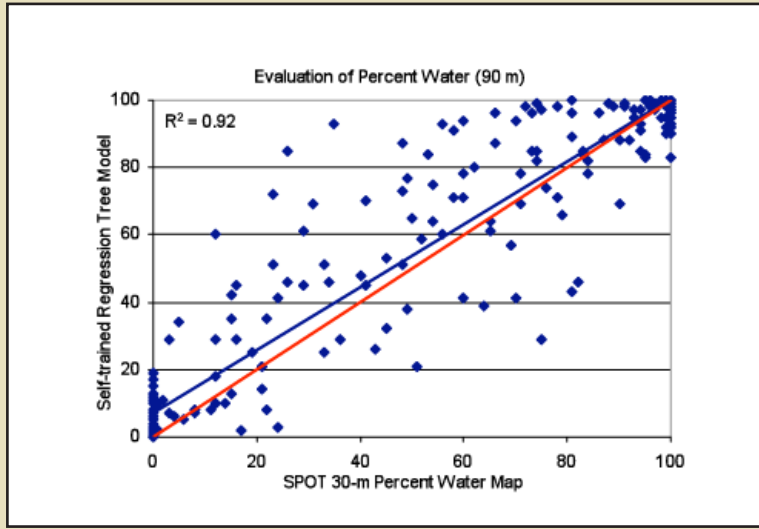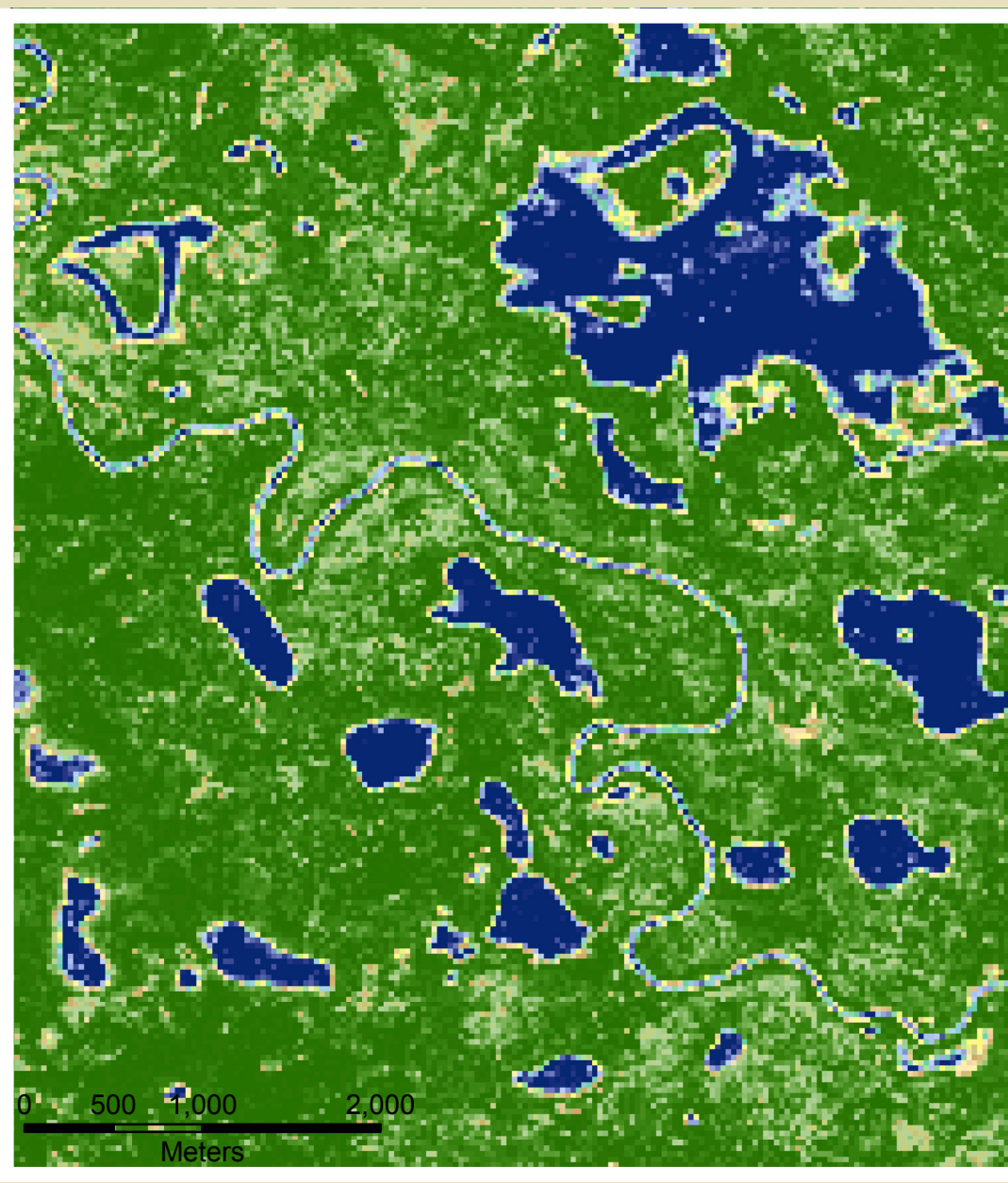
A Landsat subpixel percent water model, predicted directly from a 5-m SPOT image from the same time period, provided an ideal mean for evaluating our new technique. The modeling process included converting the 5-m water/nonwater image to a 30-m percent water map. The 30-m SPOT percent water map was then used to develop a Landsat percent water regression model. The regression model was then applied to the entire Landsat scene.
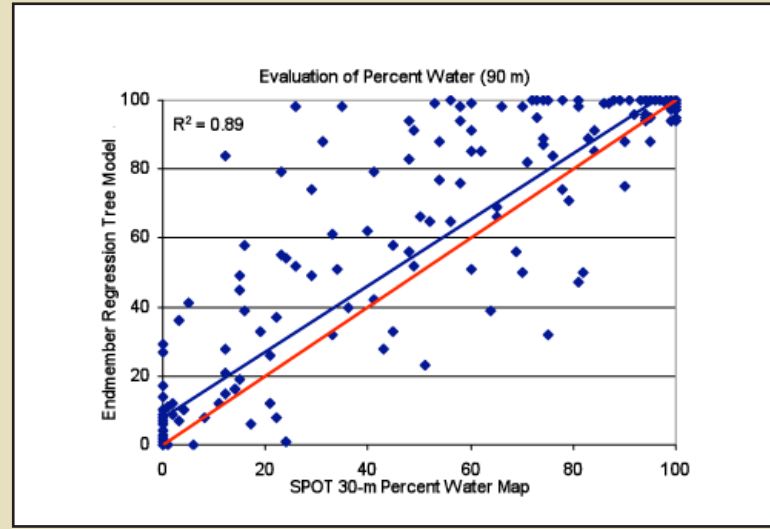
**SPOT Regression Tree Model (MSE: 11.15)**
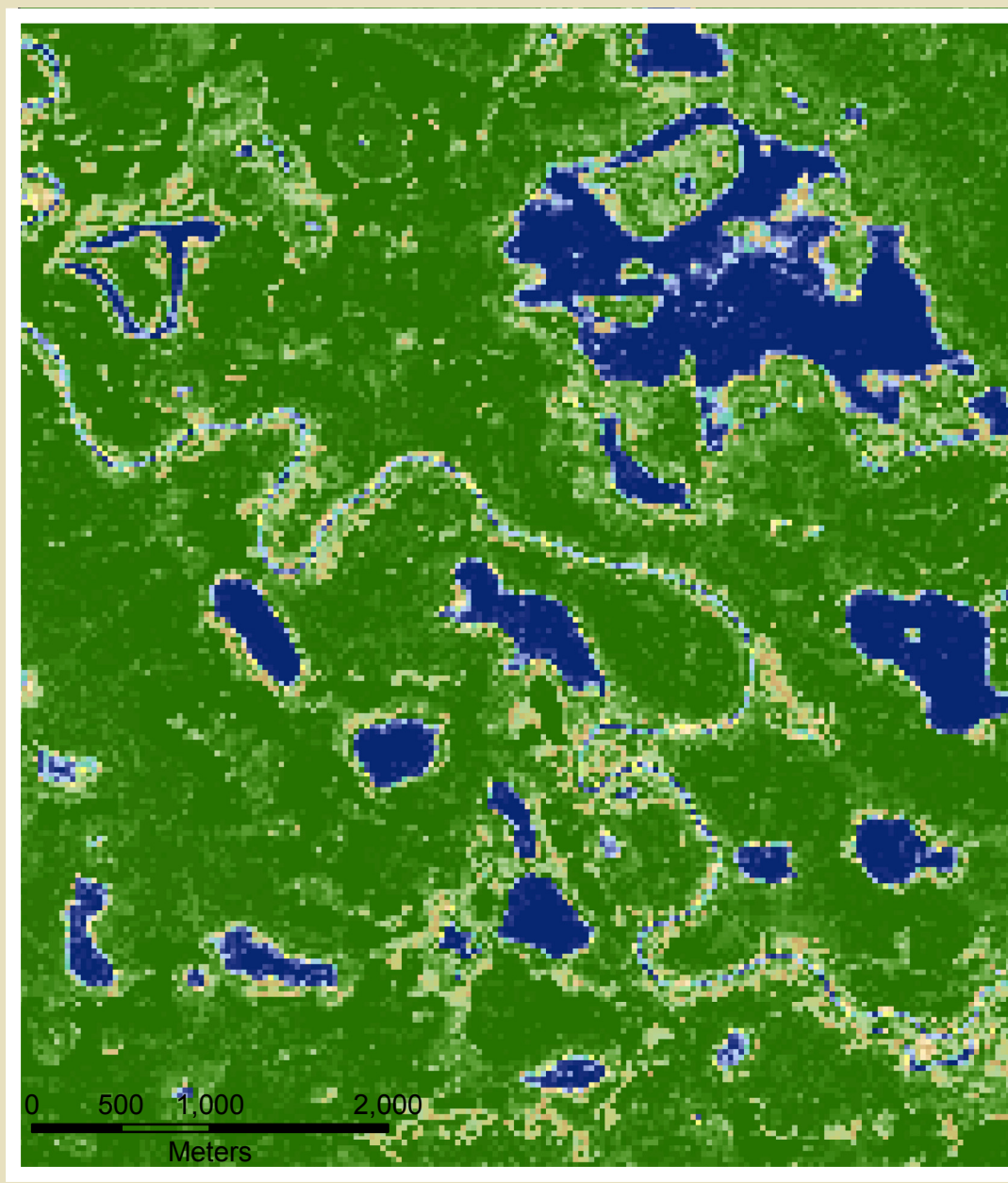
The endmember approach for constructing a rule-based percent water model restricts the training data to Landsat pixels that contained 0 and 100% water as determined from visual interpretation of the Landsat image. A regression tree model was used to predict continuous fractions (percents) of water in all pixels. This technique assumed that the spectral properties of both endmembers were representative of spectral percentages (fractions) that fall between the minimum (0) and maximum (100) percent.

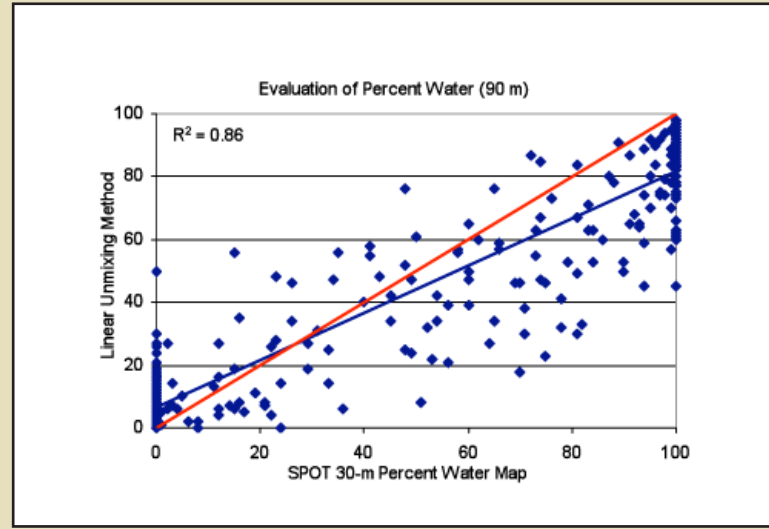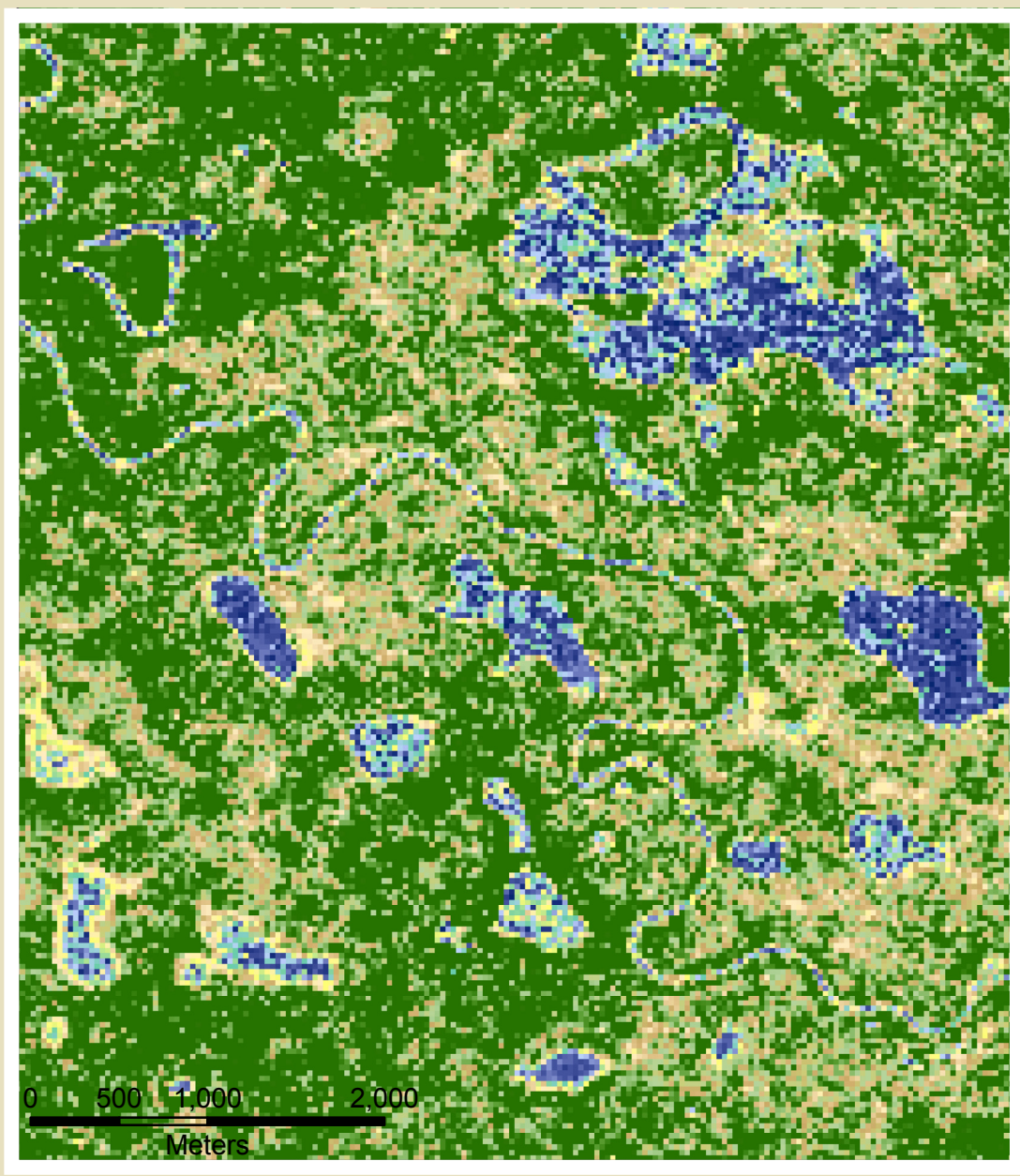**Endmember Regression Tree Model (MSE: 15.65)**
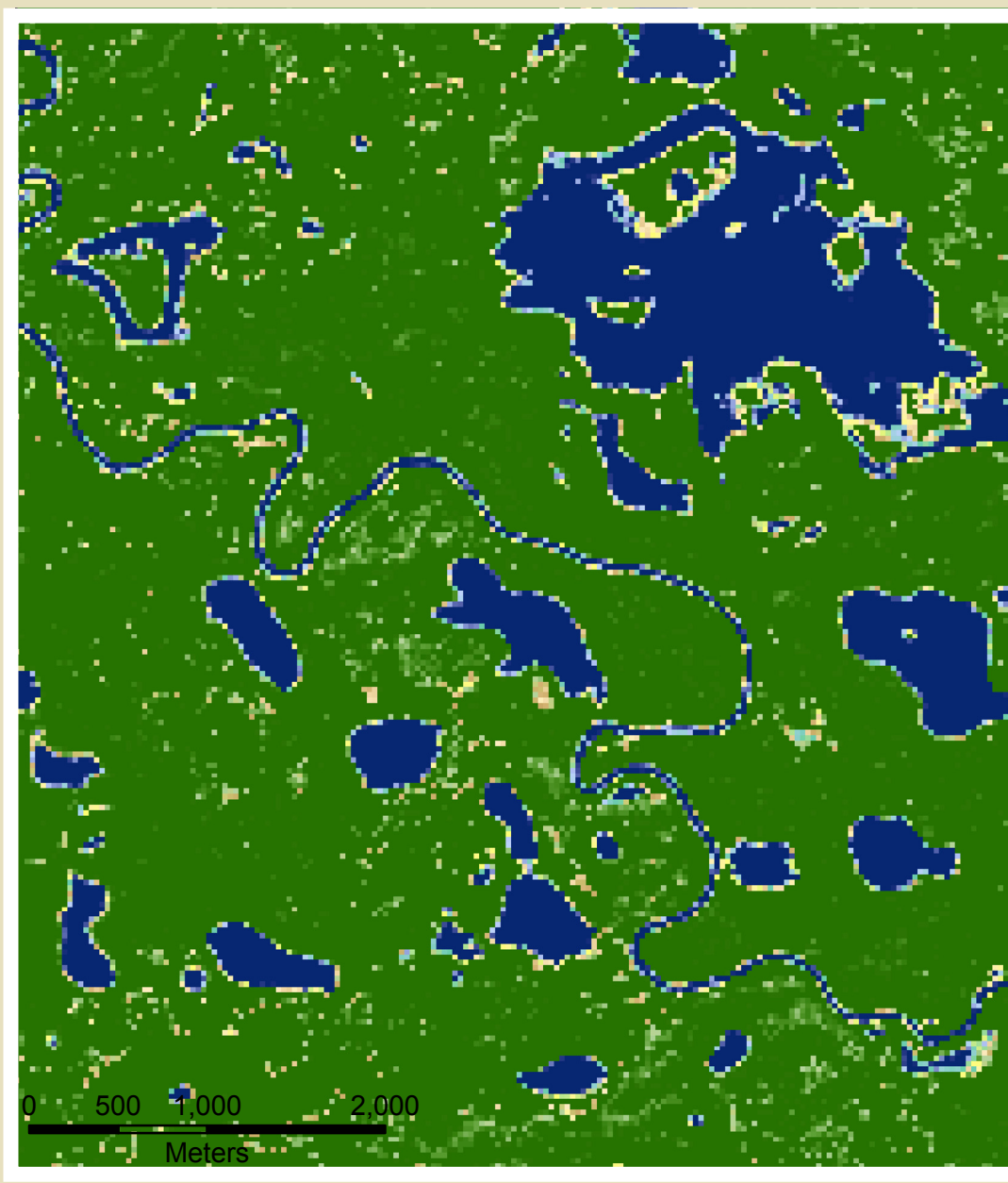
The linear unmixing approach also decomposes fractional components of pixels. This unmixing technique used a spectral library consisting of each endmember's spectra, limited to pure pixels containing water, vegetation, bare soil, and shadows in the Landsat image. Representative pure pixels establish the reflectance signature of the endmembers. An unconstrained unmixing algorithm (ENVI, 1999) was applied to the Landsat image to estimate the fractions (percent) of each endmember within a Landsat pixel. The sum of each pixel's fractions was then constrained to 100%.
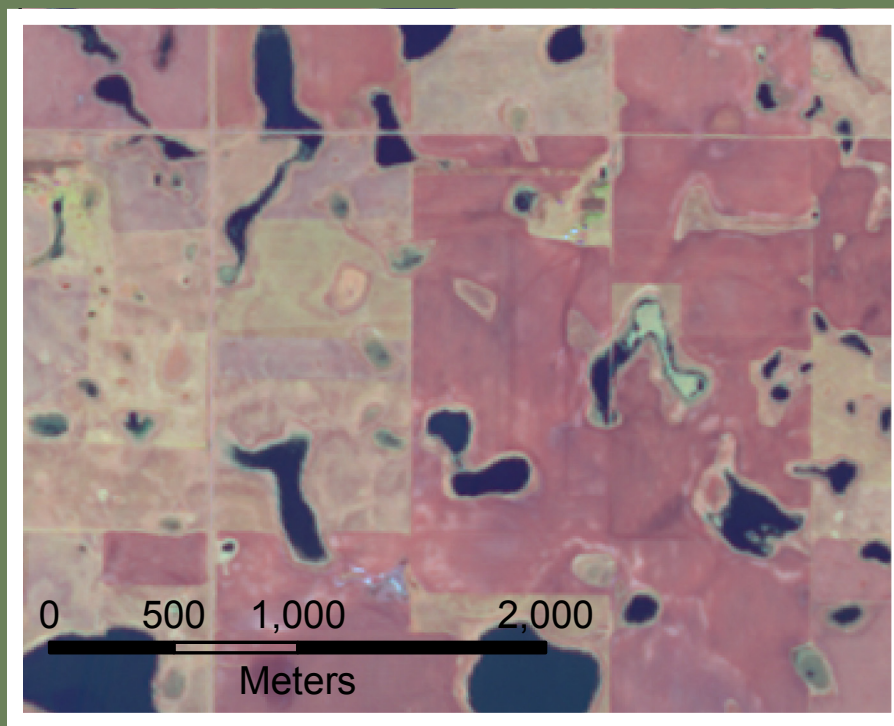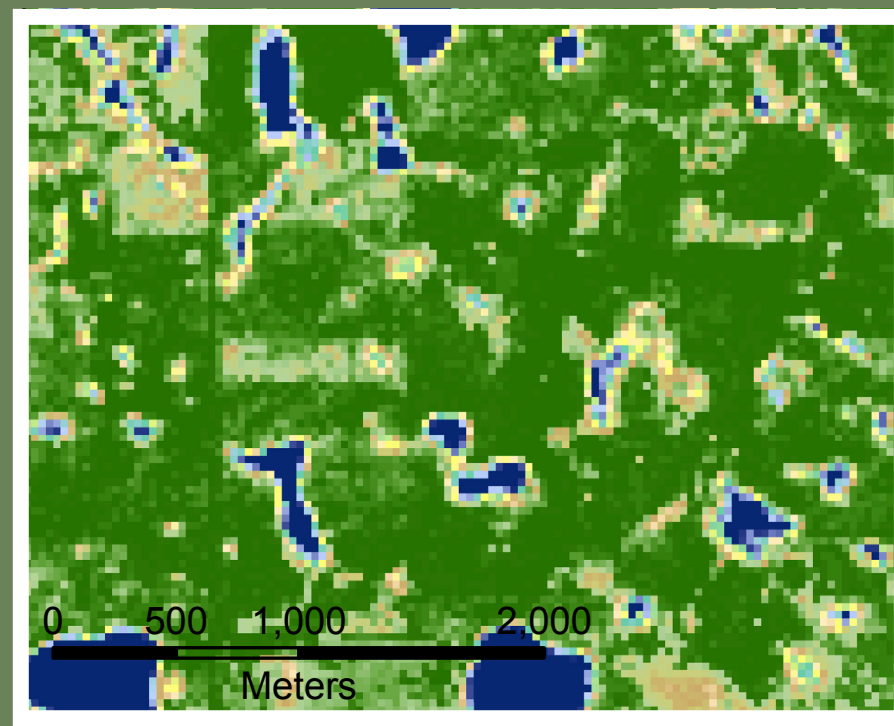
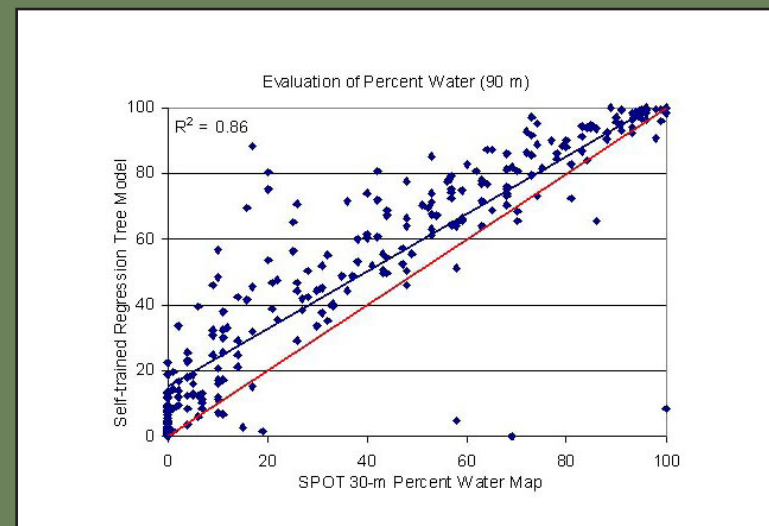**Linear Unmixing Method (MSE: 17.88)**

SPOT 5 (RGB = 4,1,3) image of the PPR

Self-trained Regression Tree Model

Self-trained Regression Tree Model (MSE: 16.57)

Discussion

The 90-m regression tree method, when applied in the PPR, produced better results than in the Yukon Flats upon visual inspection. The lack of tree shadows increased our ability to correctly identify water bodies, generally decreasing the number of incorrectly identified water pixels. In both the Yukon Flats and PPR 90-m regression models, the R^2 were 0.96 and 0.86, respectively, and the mean absolute errors were 5.2% and 5.3%, respectively, showing high goodness-of-fit. In the Yukon Flats, the new self-trained percent water mapping method produced a model with a higher R^2 value (0.92) than the linear unmixing classification (R^2=0.86) method.

The new method was also comparable to our control, percent water mapped directly from a high resolution SPOT image (R^2=0.93), but does not require high resolution training images for model development.

References

Downing, J.A., Y. T. Prairie, J. J. Cole, C. M. Duarte, L. J. Tranvik, R. G. Striegl, W. H. McDowell, P. Kortelainen, N. F. Caraco, J. M. Melack, & J. J. Middelburg. 2006. The global abundance and size distribution of lakes, ponds, and impoundments. *Limnol. Oceanogr.*, 51(5), 2388–2397.
ENVI 4.4. 2005. ENVI User's Guide. Washington D.C.: Research Systems, Inc.
Krankina, O.N., D. Pflugmacher, M. Friedl, W.B. Cohen, P. Nelson, and A. Baccini. 2008. Meeting the challenge of mapping peatlands with remotely sensed data. *Biogeosciences*, 5, 1809–1820.
Quinlan, J. R. 1993. C4.5: Programs for Machine Learning. Morgan Kaufmann, Los Altos.
Quinlan, J. R. 2008. See5: An Informal Tutorial. http://www.rulequest.com. April 2009.
Quinlan, J. R. 2009. An Overview of Cubist. http://www.rulequest.com. April 2009.
U.S. Department of Agriculture. LANDFIRE, Canopy Layer.

U.S. Department of the Interior
U.S. Geological Survey

May 2009